

LMMpro

User's Manual

version 1.xx

The Langmuir Optimization Program
plus
The Michaelis-Menten Optimization Program

Cristian P. Schulthess

LMMpro User's Manual

A tutorial on the use of LMMpro, version 1.xx.

Cristian P. Schulthess

Schulthess, C.P. 2007. LMMpro user's manual, version 1.xx. Alfisol, LLC, Coventry, CT.

Copyright © 2007 by Cristian P. Schulthess.

All Rights Reserved (including those of translation into other languages). No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the copyright holder.

Reproduction of this material or portions thereof is allowed for teaching purposes only, and also only if the name of the copyright holder and www.alfisol.com are clearly shown. This permission does not apply to the posting of this material or portion thereof in any electronic format, even if it is for teaching purposes.

The use of product or corporate names in this book may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

Alfisol, LLC

This material is distributed by Alfisol, LLC, of Coventry, Connecticut, USA. The objectives of this company are to promote products for soil and environmental scientists, to attract the general public toward soil and environmental sciences, and to help young scientists market their products.

This manual and other products are available on-line at www.alfisol.com.

Preface

The primary purpose of the LMMpro program is to find the best parameter estimates for the Langmuir Equation and the Michaelis-Menten Equation. These equations are used extensively in a wide range of research projects involving adsorption reactions and enzyme kinetics. I hope you will find the software well suited for your research needs.

These fundamental equations were introduced by Langmuir in 1918 and by Michaelis and Menten in 1913. They are among the best known equations in the biological sciences and environmental sciences. Based on my teaching experiences, I have found that students find these equations somewhat easy to understand. The difficulty, however, arises with learning how to work with the mathematics of data manipulation, the calculation of linear and nonlinear regressions, and the interpretation of the results. I hope that this software will help instructors with this topic in their lectures and lab exercises. Similarly, I hope that students will enjoy the fact that this software removes much of the tedium from their work load, which in turn will further their ability to learn how to apply these famous equations.

The LMMpro software was written in C#. With over 17,500 lines of code, I think you will find it rather user friendly and a very big improvement from whatever else you've been using for these equations up to now. In kind consideration for your fellow colleagues that are also using LMMpro, please try to promptly tell us if you find any bugs in the program. The software comes with an easy link for "report-a-bug" and another easy link for "check-for-update". I thank you in advance for your comments. Version 1.00 was released on 12 October 2007. Please visit www.alfisol.com for additional information about LMMpro.

In closing, I thank Samir A. Mohamed for his many long hours of work on the code. He taught me a great deal about C# and helped me with the program's code. I would not have completed this project in a timely manner without his help.

Cristian P. Schulthess
October 2007

Table of Contents

Tutorial: Overview of Theory	1
What does LMMpro do?	2
About the LMMpro name	2
The Langmuir Equation — an overview	3
The Michaelis-Menten Equation — an overview	5
History of Linear Regression Methods for Langmuir & Michaelis-Menten Equations	7
Linear Regressions	9
Deriving Linear Regression Methods	12
NLLS Regressions	15
v-NLLS Optimization	17
n-NLLS Optimization	20
Axes Conversion Factor (ACF)	23
Correlation Coefficient (r^2)	25
Correlation Coefficient (η^2)	26
Correlation Coefficient (η_*^2)	27
Do's & Don't's	28
 Tutorial: Overview of Software	 30
Demo Mode versus Full Access Mode	31
How to Register the Software	32
How to Check for Software Updates	33
Options	34
Run Demo	36
Start Optimization Program (O.P.)	38
Data Entry	39
Regression Execution Errors	41
Graph Commands	43

Welcome to LMMpro!

Tutorial: Overview of Theory

Select a topic:

- [What does LMMpro do?](#)
- [Langmuir Equation Overview](#)
- [Michaelis-Menten Equation Overview](#)
- [History of Linear Regression Methods](#)
- [Linear Regressions](#)
- [Deriving Linear Regression Methods](#)
- [NLLS Regressions](#)
- [v-NLLS Optimization](#)
- [n-NLLS Optimization](#)
- [Axes Conversion Factor \(ACF\)](#)
- [Correlation Coefficient \(\$r^2\$ \)](#)
- [Correlation Coefficient \(\$\eta^2\$ \)](#)
- [Correlation Coefficient \(\$\eta_*^2\$ \)](#)
- [Do's & Don't's](#)

-
- [Tutorial: Overview of Software](#)

<Reference = CTT-0>

What does LMMpro do?

This program will help you optimize the equation parameters for the **Langmuir Equation** and the **Michaelis-Menten Equation**. It uses seven different optimization methods.

As a **teaching tool**, you will be able to:

- Explain how the plots are generated.
- Show that the mathematics of analysis is easy to use if the data and theory are perfect.
- Show that the mathematics of analysis is sensitive to data error, resulting in biased results.
- Show that the mathematics of analysis is sensitive to regression method choice, resulting in biased results.
- Show that the mathematics of analysis is sensitive to theory error, resulting in misleading conclusions.

As a **research tool**, you will be able to get the best values possible for your data. You will also find it easier to evaluate if your data and theory are reasonable (see errors noted above).

Please note that every experiment is unique. LMMpro does not identify specific errors in either the data set used, the regression method used, or the theory applied. However, LMMpro will usually help you determine if everything is looking good or if something somewhere is slightly off. With proper use, you might extract some hints about where the problems (if any) are located.

About the LMMpro name:

"L" is for Langmuir, "MM" is for Michaelis-Menten, and "pro" is for program. Fortunately, "pro" is also for professional, and that is totally acceptable.

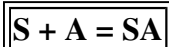
The Langmuir Equation --- an overview.

The Langmuir Equation was introduced by Irwin Langmuir in 1918. It was originally formulated for predicting the sorption of gases onto a solid phase. It is now also used to predict the sorption of aqueous compounds onto a solid phase. This is a mechanistic model that assumes that one reaction is involved in the sorption process, and that the distribution of the compounds between the two phases (gas-solid or liquid-solid) is controlled by an equilibrium constant. Langmuir received the Nobel Prize in Chemistry in 1932 for his work.

The Langmuir Equation assumes that the adsorption reaction involves a single reaction with a constant energy of adsorption. Using mass action laws for its derivation, these are the only assumptions made. Because of its simplicity, it is often the first theory tested on any adsorption data set.

Properly optimizing the parameters involved (namely $\Gamma_{\max}$ and K) will greatly impact the conclusions made about the data collected. It may also impact the decisions made by the researcher as to which objective the current project should pursue next. The equation can be derived as follows:

- Assume that only one reaction is involved, expressed as:



where **S** is the solid surface with nothing on it, **A** is the compound in the liquid or gaseous phase, and **SA** is the solid surface with compound **A** on it.

- Let $c = (\text{A})$ = equilibrium concentration of compound **A** in the liquid or gaseous phase, and define the equilibrium constant, **K**, as products over reactants:

$$\boxed{K = \frac{\{\text{SA}\}}{\{\text{S}\} c}}$$

- Define the total number of reactive sites, $\Gamma_{\max}$:

$$\boxed{\Gamma_{\max} = \{\text{SA}\} + \{\text{S}\}}$$

- Combine the equations above in adsorption terms; that is, using $\{\text{SA}\}$ terms:

$$\boxed{\Gamma_{\max} = \{\text{SA}\} + \frac{\{\text{SA}\}}{K c}}$$

- Finally, solve for {SA}, and let $\Gamma = \{SA\}$. The result is the Langmuir equation:

$$\Gamma = \frac{\Gamma_{\max} K c}{1 + K c}$$

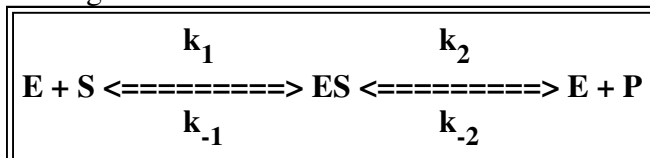
If K is known, if $\Gamma_{\max}$ is known, and if the reaction is as described above, then the amount adsorbed (Γ) can be predicted for any concentration (c) of compound **A**. The two parameters ($\Gamma_{\max}$ and K) are obtained by regression analysis of known c and Γ values.

<Reference = CTT-2>

The Michaelis-Menten Equation --- an overview.

The Michaelis-Menten Equation was introduced in 1913. It correlates the initial enzyme reaction rate (v) with the amount of substrate present (S). The equation can be derived as follows:

- Let the enzyme (E) combine with the substrate to form the enzyme-substrate complex (ES), which then proceeds to form the product (P). The enzyme (E) is recycled to be used again:



where k_i values are the reaction rate constants.

- The rate of production of product (P) is:

$$\text{rate} = d[P]/dt = v = k_2[ES] - k_{-2}[E][P] \quad \text{..... Equation [1]}$$

which can be simplified since we seek to describe the *initial* reaction rate only. That is, the initial condition is $[P] = 0$. Note, that brackets $[i]$ denote concentration of species i . Accordingly, we now have:

$$v = k_2[ES] \quad \text{..... Equation [2]}$$

- The maximum rate possible will occur when the maximum amount of enzyme exists in the ES complex form. From mass balance principles, we know that the total amount of enzyme (E_T) present is the sum of the uncomplexed form (E) plus the complexed form (ES):

$$[E]_T = [E] + [ES] \quad \text{..... Equation [3]}$$

Accordingly, the maximum rate ($V_{\max}$) of production of the product (P) occurs when $[ES] = [E]_T$. Substituting this into Equation [2], we now have:

$$V_{\max} = k_2[E]_T \quad \text{..... Equation [4]}$$

- Similarly, the rate of production of the ES complex is:

$$d[ES]/dt = k_1[E][S] - k_{-1}[ES] - k_2[ES] + k_{-2}[E][P] \quad \text{..... Equation [5]}$$

At the start of the process, there is no product present, $[P] = 0$, and this allows us to simplify Equation [5] when elapsed time is near zero:

$$d[ES]/dt = k_1[E][S] - k_{-1}[ES] - k_2[ES] \quad \text{..... Equation [6]}$$

- It is reasonable to assume that *steady state conditions* are immediately established. Here, this means that the rate of production of ES is equal to the rate of decomposition of ES . Mathematically, this means $d[ES]/dt = 0$, where no change in concentration of

ES is observed as a function of time. If we plug this into Equation [6] and rearrange the equation, we get:

$\frac{[ES]}{[E]} = \frac{k_1[S]}{k_{-1} + k_2} = \frac{[S]}{K_M}$	 Equation [7]
--	--------------------

where

$K_M = \frac{k_{-1} + k_2}{k_1}$

and K_M = the Michaelis-Menten constant.

- Typically, we do not know the concentrations [E] or [ES], or, at least, these values are far too difficult to determine. So the final mathematical steps shown here seek to present an elegant expression for the reaction involved using measureable parameters. We continue with the ratio of Equations [2] and [4], and substitute with Equations [3] and [7] as needed:

$\frac{V_{\max}}{v} = \frac{k_2[E]_T}{k_2[ES]} = \frac{[E]_T}{[ES]} = \frac{[E] + [ES]}{[ES]} = \frac{[E]}{[ES]} + 1 = \frac{[S]}{K_M} + 1$	 Equation [8]
---	--------------------

- Rearranging Equation [8]:

$v = \frac{V_{\max} [S]}{K_M + [S]}$	 Equation [9]
--------------------------------------	--------------------

which is known as the Michaelis-Menten Equation.

If K_M is known, if $V_{\max}$ is known, and if the reaction is as described above, then the reaction rate (v) can be predicted for any initial substrate concentration (S). The two parameters ($V_{\max}$ and K_M) are obtained by regression analysis of known S and v values. It is worth noting that when $S = K_M$, the observed reaction rate (v) will equal $V_{\max} / 2$.

History of Linear Regression Methods for Langmuir & Michaelis-Menten Equations.

The Langmuir and Michaelis-Menten Equations are very similar to each other in form. They are both parabolic equations. Accordingly, the methods developed to solve these equations are also similar. The objective is to find the best K and $\Gamma_{\max}$ values of the Langmuir Equation, or the best K_M and $V_{\max}$ values of the Michaelis-Menten Equation, so that the parabolic equation fits the data with minimal error.

There are several methods available. Without computers, it is best to convert the parabolic equation into a linear form, which can then be solved by linear regression techniques. Langmuir proposed a linearization method in 1918 (known as the Langmuir linear regression), but it was not known to those using the Michaelis-Menten Equation. Accordingly, Hanes and later Woolf proposed a similar regression technique in 1932 (known as the Hanes-Woolf linear regression). Note that the Langmuir linear regression and the Hanes-Woolf linear regression are identical in form.

A common method used to solve the Michaelis-Menten Equation is the Lineweaver-Burk linear regression. Since several early environmental scientists did not know about Langmuir's linearization methods, this Lineweaver-Burk linear regression was also used extensively to solve the Langmuir Equation. This method is very sensitive to data error, and its use is now strongly discouraged.

A few years later, Eadie-Hofstee (1942, 1952) proposed another method for solving the Michaelis-Menten Equation. The Eadie-Hofstee linear regression was a significant improvement from the earlier Lineweaver-Burk linear regression, but it is still somewhat sensitive to data error. Once again, several early environmental scientists did not know about Langmuir's linearization method, and this Eadie-Hofstee linear regression was also used extensively to solve the Langmuir Equation.

Another effort to come up with a better Michaelis-Menten optimization method was presented by Scatchard in 1949. It still suffers from some sensitivity to data error. Once again, several early environmental scientists did not know about Langmuir's linearization method, and this Scatchard linear regression was also used extensively to solve the Langmuir Equation.

Finally, the best Michaelis-Menten optimization method using linear regression techniques was presented by Hanes 1932, and referred to as the Hanes-Woolf regression by Haldane (1957). As noted earlier, this technique is essentially identical to the linear regression technique presented by Langmuir in 1918. It is the least sensitive to data error. Once again, several early environmental scientists did not know about Langmuir's linearization methods, and this Hanes-Woolf linear regression was also used extensively to solve the Langmuir Equation. That is, the technique was incorrectly named the Hanes-Woolf linear regression (rather than the Langmuir linear regression) when used to solve the Langmuir Equation.

There is one more linear regression that is worth noting here. The data can be presented in a log-log form, resulting in a linear pattern with a slope = 1.0. It is optimized via an iteration process, and, hence, this method is not used. That is, this method is not easy to use without computers.

<Reference = CTT-4>

Linear Regressions

It is rather easy to optimize a parabolic equation. The linear regression methods used by LMMpro are tabulated below. The year each method was developed is also shown along with the names of the authors of these regressions methods. Note that if we allow $K = 1/K_M$, then the form of these two parabolic equations is essentially the same.

Name of Linear Regression:	Langmuir Equation (1918): $\Gamma = \frac{\Gamma_{\max} K c}{1 + K c}$ Γ = Amount adsorbed $\Gamma_{\max}$ = Maximum adsorption quantity K = Reaction equilibrium constant c = Aqueous Equilibrium concentration	Michaelis-Menten Equation (1913): $v = \frac{V_{\max} S}{K_M + S}$ v = Overall rate of reaction $V_{\max}$ = Maximum reaction rate K_M = Michaelis-Menten constant S = Substrate concentration	Comments: Note: the term " <i>original graph</i> " in the comments below refers to graph generated by either of the two parabolic equations expressed here, namely the Langmuir Equation or the Michaelis-Menten Equation, plus the (c, Γ) or the (S, v) original data.
Lineweaver-Burk (1934):	$\frac{1}{\Gamma} = \frac{1}{\Gamma_{\max}} + \frac{1}{\Gamma_{\max} K c}$ plot (1/ Γ) versus (1/c) slope = $1/(\Gamma_{\max} K)$ intercept = $1 / \Gamma_{\max}$	$\frac{1}{v} = \frac{1}{V_{\max}} + \frac{K_M}{V_{\max} S}$ plot (1/v) versus (1/S) slope = $K_M / V_{\max}$ intercept = $1 / V_{\max}$	This regression method is extremely sensitive to data error. It has a very strong bias for closely tracking the data in the lower left corner of the original graph.

Eadie-Hofstee (1942, 1952):	$\Gamma = \Gamma_{\max} - \frac{\Gamma}{K} c$ plot (Γ) versus (Γ/c) slope = $-1/K$ intercept = $\Gamma_{\max}$	$v = V_{\max} - \frac{vK_M}{S}$ plot (v) versus (v/S) slope = $-K_M$ intercept = $V_{\max}$	This regression method has some sensitivity to data error. It has some bias for closely tracking the data in the lower left corner of the original graph. Note that if you invert the x,y-axes, then this would convert into the Scatchard regression. A Scatchard regression is equivalent to a regression that minimizes the <i>horizontal</i> error in an Eadie-Hofstee plot.
Scatchard (1949):	$\frac{\Gamma}{c} = K \Gamma_{\max} - K \Gamma$ plot (Γ/c) versus (Γ) slope = $-K$ intercept = $K \Gamma_{\max}$	$\frac{v}{S} = \frac{V_{\max}}{K_M} - \frac{v}{K_M}$ plot (v/S) versus (v) slope = $-1/K_M$ intercept = $V_{\max} / K_M$	This regression method has some sensitivity to data error. It has some bias for closely tracking the data in the upper right corner of the original graph. Note that if you invert the x,y-axes, then this would convert into the Eadie-Hofstee regression. An Eadie-Hofstee regression is equivalent to a regression that minimizes the <i>horizontal</i> error in a Scatchard plot.

left: Langmuir (1918) right: Hanes- Woolf (1932, 1957):	$\frac{c}{\Gamma} = \frac{c}{\Gamma_{\max}} + \frac{1}{K \Gamma_{\max}}$ plot (c/Γ) versus (c) slope = 1 / Γ_{max} intercept = 1/(K Γ_{max})	$\frac{S}{v} = \frac{S}{V_{\max}} + \frac{K_M}{V_{\max}}$ plot (S/v) versus (S) slope = -1 / V_{max} intercept = K_M / V_{max}	This regression method has very little sensitivity to data error. It has some bias for closely tracking the data in the middle portion of the graph plus the upper right corner of the original graph. This linear regression technique was first presented by Langmuir in 1918. Although he received the Nobel Prize in 1932, the method he used to optimize the parabolic equation was apparently totally missed by others. It was later presented by Hanes (1932) and referred to as the Hanes-Woolf regression by Haldane (1957), and this regression method often carries their names. This software (LMMpro) refers to this regression method as the "Langmuir Linear Regression Method" when used to solve the Langmuir adsorption isotherm.
log-log:	$\log \frac{\theta}{1 - \theta} = \log c + \log K$ plot (log [θ/(1-θ)]) versus (log c) θ = Γ / Γ_{max} slope = 1 intercept = log K	$\log \frac{\theta}{1 - \theta} = \log S - \log K_M$ plot (log [θ/(1-θ)]) versus (log S) θ = v / V_{max} slope = 1 intercept = -log K_M	This regression method has very little sensitivity to data error. This is the only linear regression method that is too difficult to solve by hand. It must be solved via an iterative loop that finds the equation's best maximum value (and, hence, the best θ value). The best θ value is the one with the smallest linear regression error. Note that the slope is fixed and it is set equal to 1.0.

Note that all linear and nonlinear regression methods are also sensitive to theory error. That is, a small deviation in the data from the Langmuir theory predictions or Michaelis-Menten theory predictions is not necessarily an expression of an error in the data collected. It may instead be due to a slightly incomplete mathematical expression of the true nature of the process involved.

<Reference = CTT-5>

Deriving Linear Regression Methods

Given a line $y = mx + b$,

how do we determine what the best line's slope (m) and intercept (b) should be for a given set of data (x,y)?

We could randomly guess the m and b values of the line, tabulate the error of each, and then identify which guess results in the least error. This can be done with a computer, but due to the lack of computers in the late 1700's, an alternate method was proposed that is much better. The method is called a least squares linear regression method because we solve the problem backwards (hence the term "regression"); that is, we first minimize the error and we then find out which line is it that gives us that low error value. The method does not really care what the minimum error value actually is. Instead, we note that the derivative of the error function reaches a minimum and its slope is equal to zero when the error value is also at a minimum.

1. Let $\epsilon^2 = \text{error}^2 = \sum (y_i - y_p)^2$,
 where y_p = predicted value when $x = x_i$, and y_i = experimental value when $x = x_i$. Note that we work with the square of the error because it eliminates the problem of working with the absolute value of the error. Although the error can be either positive or negative, its square is always positive. It is perfectly reasonable to work with other exponents (say, 4 or 6), but using the power of 2 is easier to solve. This also explains the origin of the name of the method: it's a *least squares* linear regression.
2. Let $y_p = m x_i + b$.
3. Combine (1) and (2), and expand:

$$\epsilon^2 = \sum (y_i - m x_i - b)^2$$

$$\epsilon^2 = \sum (m^2 x_i^2 + 2mbx_i - 2mx_i y_i + b^2 - 2by_i + y_i^2)$$
4. Step 3 above yields a plot of ϵ^2 as a function of the line's slope m and the line's intercept b. As we approach the best value for m and b, ϵ^2 approaches its minimum value. If we overshoot the m and b values, the ϵ^2 value begins to increase again. If we stop right at the bottom of the plot, right where the best m and b values exist, then the ϵ^2 is at its minimum and the tangent of this error function right at that point is zero. The tangent at any point in the curve is the first-derivative of the curve. We seek the point where the first-derivative equals zero, and where the error as a function of m and b is at a minimum:

$$\frac{d\epsilon^2}{dm} = 0, \text{ and } \frac{d\epsilon^2}{db} = 0$$

5. Solve Equation [3] for the minimum as a function of m:

$$\frac{d\varepsilon^2}{dm} = 0 = 2m\Sigma x_i^2 - 2b\Sigma x_i - 2\Sigma(x_i y_i)$$

6. Solve Equation [3] for the minimum as a function of b:

$$\frac{d\varepsilon^2}{db} = 0 = 2m\Sigma x_i + 2\Sigma b - 2\Sigma y_i$$

7. To simplify the notation (for number of points = n):

$$\text{Let } S_x = \Sigma x_i$$

$$\text{Let } S_y = \Sigma y_i$$

$$\text{Let } S_{xy} = \Sigma (x_i y_i)$$

$$\text{Let } S_{xx} = \Sigma (x_i^2)$$

$$\text{Let } nb = \Sigma b$$

8. From step 5, solve for m:

$$m = \frac{b S_x + S_{xy}}{S_{xx}}$$

If the intercept's value b is fixed, then all parameters are known, and solve for m.

9. From step 6, solve for b:

$$b = (S_y - m S_x) / n .$$

If the slope's value m is fixed, then all parameters are known, and solve for b.

10. If neither the slope nor the intercept are known, then combine step 9 and 8, and solve for m:

$$0 = m S_{xx} - S_x (S_y - m S_x)/n - S_{xy}$$

$$0 = m S_{xx} - S_x S_y/n + m S_x S_x/n - S_{xy}$$

$$m (n S_x + S_{xx}) = S_{xy} + S_x S_y$$

$$m = \frac{n S_{xy} + S_x S_y}{n S_x + S_{xx}}$$

Now plug this value of m into step 9 to get the value of b.

Note that data will be algebraically balanced above and below the regression line. This remains true even if the m or b values (but not both) are fixed by the user.

This regression analysis is easy enough to do by hand. It is even easier today with computers that do it for us. Note that his method was discovered by Carl Friedrich Gauss in 1794. He was 17 years old at the time. The method was published by Adrien-Marie Legendre in 1806.

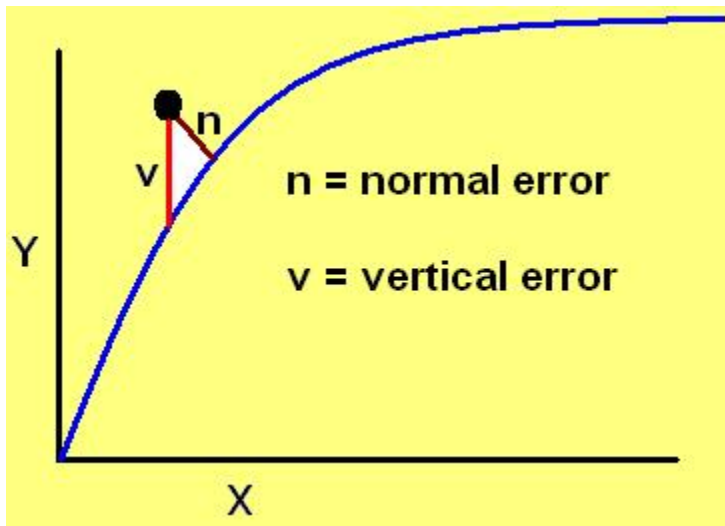
<Reference = CTT-6>

NLLS Regressions

With computers, it is rather easy to optimize a parabolic equation using nonlinear methods. The optimization minimizes the squares of the error terms, and hence these methods are known as Nonlinear Least Squares (NLLS) methods. There are two NLLS methods used by LMMpro: one that minimizes the vertical error (v-NLLS), the other minimizes the normal error (n-NLLS). The equation parameters are incrementally changed until a minimum in the sum-of-the-squares-of-the-errors is attained.

Using **v-NLLS**, the error is defined as the square of the *vertical* distance between each datum point and the predicted curve.

Using **n-NLLS**, the error is defined as the square of the *normal* distance between each datum point and the predicted curve. The normal distance is defined as the perpendicular distance from the nearest place on the curve.



When using n-NLLS, the slope of the error line between the datum point and the predicted curve will change depending on the location of the datum point. It will be nearly horizontal in the lower left corner of the graph. It will be nearly vertical in the upper right corner of the graph. Accordingly, with n-NLLS you must decide upon the relative importance of a horizontal error versus a vertical error. A diagonal error line will have both vertical error and horizontal error components. The ACF value is used to determine the relative impact of these types of error.

The **Axes Conversion Factor (ACF)** describes the relative importance (weight) of an error in x versus an error in y . IF $ACF = 1$, then x -axis and y -axis values have the same weight. A large ACF value (>1) means that the y -axis values are weighed more strongly. A small ACF value (<1) means that the x -axis values are weighed more strongly.

NLLS Regression Bias

The v-NLLS method is biased toward optimizing the fit with the data in the lower left side of the graph. This is because this area of the graph will generate large errors if the predicted curve is not properly tracking these data. Notice that even a small change in the predicted curve will slightly shift the location of the sharp rise in the parabolic equation, which, in turn, results in a large vertical error with each datum point in that area.

Using the n-NLLS method, a small change in the predicted curve will not result in any significant increase in its distance to any of the data points. Accordingly, the n-NLLS method is not biased toward optimizing the fit with the data in any particular area of the graph.

The n-NLLS regression method used by LMMpro was presented by Schulthess & Dey in 1996 (*Soil Sci. Soc. Am. J.* 60:433-442).

<Reference = CTT-7>

v-NLLS Optimization

The v-NLLS regression method used by LMMpro is based on an adaptation of the optimization method discussed by Persoff & Thomas in 1988 (*Soil Sci. Soc. Am. J.* 52:886-889).

v-NLLS stands for vertical nonlinear least squares. This regression method optimizes the parameters of the equation without converting the equation into another form or shape. The best fit is that equation that yields the smallest error. The error of each datum point is defined as the vertical difference between the predicted value and the actual value. The predicted value is the value calculated for a given x_i , and the actual value is the value measured for a given x_i .

For the Langmuir Equation, the optimization is as follows:

1. Let $\Sigma \varepsilon_i^2 = \Sigma \text{error}^2 = \Sigma (\text{measured} - \text{predicted})^2$
2. Substitute the Langmuir Equation for predicted values:

$$\Sigma \varepsilon_i^2 = \Sigma \left[\Gamma_i - \frac{\Gamma_{\max} K c_i}{1 + K c_i} \right]^2$$

3. Expand the square term:

$$\Sigma \varepsilon_i^2 = \Sigma \left[\frac{\Gamma_{\max} K c_i}{1 + K c_i} \right]^2 - \Sigma \left[\frac{2 \Gamma_i \Gamma_{\max} K c_i}{1 + K c_i} \right] + \Sigma \Gamma_i^2$$

4. To optimize $\Gamma_{\max}$, set the first derivative equal to zero and solve:

$$\frac{d(\Sigma \varepsilon_i^2)}{d\Gamma_{\max}} = 0 = 2 \Gamma_{\max} \Sigma \left[\frac{K c_i}{1 + K c_i} \right]^2 - \Sigma \left[\frac{2 \Gamma_i K c_i}{1 + K c_i} \right] + 0$$

5. Solving Equation [4] for $\Gamma_{\max}$ yields:

$$\Gamma_{\max} = \frac{\sum \left[\frac{\Gamma_i K c_i}{1 + K c_i} \right]}{\sum \left[\frac{K c_i}{1 + K c_i} \right]^2}$$

6. If K is known (or fixed by the user), then use Equation [5] above to get the optimized $\Gamma_{\max}$ value.
7. If K is not known, guess K values, and then use Equation [5] to get the corresponding best $\Gamma_{\max}$ value. Finally, use Equation [2] to determine the error that corresponds to the guesses made. Repeat the process by incrementally increasing or decreasing the K value guesses until the minimum error condition is found.
8. If $\Gamma_{\max}$ is known (or fixed by the user), then Equation [5] is not needed. The best value of K , however, is still determined with Equation [2] and an iteration loop to find the condition with the minimum error.

For the Michaelis-Menten Equation, the optimization is as follows:

1. Let $\sum \varepsilon_i^2 = \sum \text{error}^2 = \sum (\text{measured} - \text{predicted})^2$
2. Substitute the Michaelis-Menten Equation for predicted values:

$$\sum \varepsilon_i^2 = \sum \left[v_i - \frac{V_{\max} S_i}{K_M + S_i} \right]^2$$

3. Expand the square term:

$$\sum \varepsilon_i^2 = \sum \left[\frac{V_{\max} S_i}{K_M + S_i} \right]^2 - \sum \left[\frac{2 v_i V_{\max} S_i}{K_M + S_i} \right] + \sum v_i^2$$

4. To optimize $V_{\max}$, set the first derivative equal to zero and solve:

$$\frac{d(\sum \varepsilon_i^2)}{dV_{\max}} = 0 = 2 V_{\max} \sum \left[\frac{S_i}{K_M + S_i} \right]^2 - \sum \left[\frac{2 v_i S_i}{K_M + S_i} \right] + 0$$

5. Solving Equation [4] for $V_{\max}$ yields:

$$V_{\max} = \frac{\sum \left[\frac{v_i S_i}{K_M + S_i} \right]}{\sum \left[\frac{S_i}{K_M + S_i} \right]^2}$$

6. If K_M is known (or fixed by the user), then use Equation [5] above to get the optimized $V_{\max}$ value.
7. If K_M is not known, guess K_M values, and then use Equation [5] to get the corresponding best $V_{\max}$ value. Finally, use Equation [2] to determine the error that corresponds to the guesses made. Repeat the process by incrementally increasing or decreasing the K_M value guesses until the minimum error condition is found.
8. If $V_{\max}$ is known (or fixed by the user), then Equation [5] is not needed. The best value of K_M , however, is still determined with Equation [2] and an iteration loop to find the condition with the minimum error.

Note that the v-NLLS regression will optimize the parameters assuming that minimizing the vertical error yields the best results. This method does have some bias in favor of any region of the curve where the vertical changes are most pronounced. In other words, for a parabolic equation such as the Langmuir Equation or the Michaelis-Menten Equation, the v-NLLS regression has some bias toward the lower left region of the graph.

Also note that the v-NLLS regression does not result in an optimized curve with the data evenly distributed above and below the curve. That is, the sum of the errors above the curve is not the same value as the sum of the errors below the curve. Least squares regressions will only balance the data around the curve if the function has a constant term. For example, in $y = f(x) + b$, b is the constant term. The Langmuir Equation and the Michaelis-Menten Equation are lacking the constant term (b).

n-NLLS Optimization

The n-NLLS regression method used by LMMpro was presented by Schulthess & Dey in 1996 (*Soil Sci. Soc. Am. J.* 60:433-442).

n-NLLS stands for normal nonlinear least squares. This regression method optimizes the parameters of the equation without converting the equation into another form or shape. The best fit is that equation that yields the smallest error. The error of each datum point is defined as the distance between the datum point and its nearest point on the parabolic curve. The nearest trajectory of a datum point to the curve is a line that is normal (that is, perpendicular) to the tangent of the nearest point on the curve.

For the Langmuir Equation, the optimization involves two loops, as follows:

1. Make an initial guess of the K and $\Gamma_{\max}$ values.
Let Q_i = normal distance of each datum point to the Langmuir isotherm predicted. The sign of Q_i is negative if the point is below or to the right of the curve.
2. Let $SUM1 = | \Sigma Q_i |$ = absolute value of the sum of the errors.
3. Let $SUM2 = \Sigma Q_i^2$.
4. Repeat continuously steps 2 and 3 using a different $\Gamma_{\max}$ value until a minimum $SUM1$ value is achieved.
5. Repeat continuously steps 2, 3, and 4 using a different K value until a minimum $SUM2$ value is achieved.

When you exit both of these loops, you will have the optimized K and $\Gamma_{\max}$ values.

For the Michaelis-Menten Equation, the optimization involves two loops, as follows:

1. Make an initial guess of the K_M and $V_{\max}$ values.
Let Q_i = normal distance of each datum point to the Michaelis-Menten reaction rate predicted. The sign of Q_i is negative if the point is below or to the right of the curve.
2. Let $SUM1 = | \Sigma Q_i |$ = absolute value of the sum of the errors.
3. Let $SUM2 = \Sigma Q_i^2$.

4. Repeat continuously steps 2 and 3 using a different $V_{\max}$ value until a minimum SUM1 value is achieved.
5. Repeat continuously steps 2, 3, and 4 using a different K_M value until a minimum SUM2 value is achieved.

When you exit both of these loops, you will have the optimized K_M and $V_{\max}$ values.

Calculating the coordinates of the closest point on the curve.

An exact value of Q_i defined above can be calculated as follows.

1. Let (m,n) = coordinates of datum point.
Let (r,s) = coordinates of closest point on the predicted curve.

2. The distance between two points is defined by:

$$Q = [(m - r)^2 + (n - s)^2]^{0.5}$$

3. Substitute the theoretical equation for s into the equation above:

For Langmuir Equation:	$s = \Gamma = \frac{\Gamma_{\max} K r}{1 + K r}$
For Michaelis-Menten Equation:	$s = v = \frac{V_{\max} r}{K_M + r}$

4. Using differential equations, we solve to minimize Q as a function of r.
That is, $dQ/dr = 0$. Simplify the result into the following equation of the fourth degree:

$$A r^4 + B r^3 + C r^2 + D r + E = 0$$

5. For the Langmuir Equation:

Let $A = K^3$.

Let $B = 3K^2 - mK^3$.

Let $C = 3K - 3mK^2$.

Let $D = 1 - 3mK + K^2\Gamma_{\max}^2 - nK^2\Gamma_{\max}$.

Let $E = -m - nK\Gamma_{\max}$.

Note: the definition of D above is correct. It is not correct in the original publication (Schulthess & Dey, 1996).

Also note that n and $\Gamma_{\max}$ are multiplied by the axes conversion factor (ACF) in order to maintain the parity of the units in each component of this equation of the fourth degree.

6. For the Michaelis-Menten Equation:

$$\text{Let } A = K_M^{-3}.$$

$$\text{Let } B = 3K_M^{-2} - mK_M^{-3}.$$

$$\text{Let } C = 3/K_M - 3mK_M^{-2}.$$

$$\text{Let } D = 1 - 3m/K_M + K_M^{-2}V_{\max}^2 - nK_M^{-2}V_{\max}.$$

$$\text{Let } E = -m - (n/K_M)V_{\max}.$$

Note that n and $V_{\max}$ are multiplied by the axes conversion factor (ACF) in order to maintain the parity of the units in each component of this equation of the fourth degree.

7. Choose the feasible root for the value of r . The solution to the equation of the fourth degree yields four roots. Only one of these is the feasible answer.
8. Once r is known, use Equation [2] and [3] to evaluate the values of s and Q .

Note that the n-NLS regression will optimize the parameters assuming that minimizing the normal error yields the best results. This method does not have any known bias in favor of any particular region of the curve. Furthermore, the ACF factor gives much flexibility to how one wishes to weigh the axes units.

Also note that the n-NLS regression will result in an optimized curve with the data evenly distributed above and below the curve. That is, the sum of the errors above the curve will be the same value as the sum of the errors below the curve. This balance is a result of the SUM1 definition outlined above. While the SUM1 definition balances the data around the curve, the SUM2 definition tightens the curve to get as close to the data as possible.

Axes Conversion Factor (ACF)

The Axes Conversion Factor (ACF) describes the relative importance (weight) of an error in x versus an error in y . If $ACF = 1$, then x axis and y axis values have the same weight. A large ACF value (>1) means that the y axis values are weighed more strongly. A small ACF value (<1) means that the x axis values are weighed more strongly.

The x -axis and y -axis values are sometimes not determined completely independent of each other. This is easy to illustrate with adsorption experiments involving a compound (A) in the liquid phase adsorbing to a solid phase. Typically, the total concentration of compound A in the mixture is known. It is then an easy calculation to determine the amount of compound A that was adsorbed by the solid phase (Γ) based on the amount of compound A remaining in the liquid phase at equilibrium (c). That is, Γ is not actually measured, but rather it is extrapolated from c . Hence, any error in the measurement of c will also express a similar error in Γ . Although there is no definitive proof for how the ACF value should be determined, we can use this relationship (or dependency of the x,y data to estimate the ACF value. An example follows that illustrates a useful procedure for calculating the ACF value.

Example for estimating the ACF value:

- Assume an adsorption experiment in a closed container. This is a batch reaction type of experiment, where the amount adsorbed by the solid is a function of the amount in solution at equilibrium.
- Assume that the data collected are such that the y -axis units are $\mu\text{g m}^{-2}$.
Assume that the data collected are such that the x -axis units are mg L^{-1} .
- The question asked here is: "one unit change in y equals how many units change in x ?"
- Assume solids concentration = 9.24 g L^{-1} .
- Assume specific surface area of the solid = $100 \text{ m}^2 \text{ g}^{-1}$.
- Given the above conditions:

$$\frac{1 \mu\text{g}}{\text{m}^2} \times \frac{100 \text{ m}^2}{\text{g}} \times \frac{9.24 \text{ g}}{\text{L}} \times \frac{\text{mg}}{1000 \mu\text{g}} = \frac{0.924 \text{ mg}}{\text{L}}$$

- For this example, $ACF = 0.924$.

What does ACF mean?

It will carry a different meaning with every application. In the adsorption example above, $ACF = 0.924$. Notice that the amount adsorbed was determined based on the amount remaining in solution. This practice is common in adsorption isotherm experiments, where the y -axis values are not truly independent of the x -axis values. A loss of 0.924 mg/L from the liquid phase is equal to an adsorption amount of $1 \mu\text{g/m}^2$. An error of 0.924 mg/L for the liquid phase concentration is equal to an adsorption error of $1 \mu\text{g/m}^2$. Arithmetically, nothing really changes if you determine the amount adsorbed by direct measurement of the solid phase because the adsorption value is directly related to the liquid phase equilibrium value as a result of the principle of mass balance. That is, the total sum of the nonadsorbed liquid portion plus the adsorbed solid portion must be constant at all times.

We can apply this example to enzyme kinetics. In enzyme kinetics studies, the rate of reaction (v) can be evaluated based on the rate of loss of substrate (S). Accordingly, an error in the measurement of S will result in an error in the measurement of v . The ACF value measures this relationship. Arithmetically, nothing really changes if you determine the reaction rate by direct measurement of the product formed because the product concentration is directly related to the reactant concentration as a result of the principle of mass balance. That is, the total sum of the product formed plus the reactant remaining must be constant at all times.

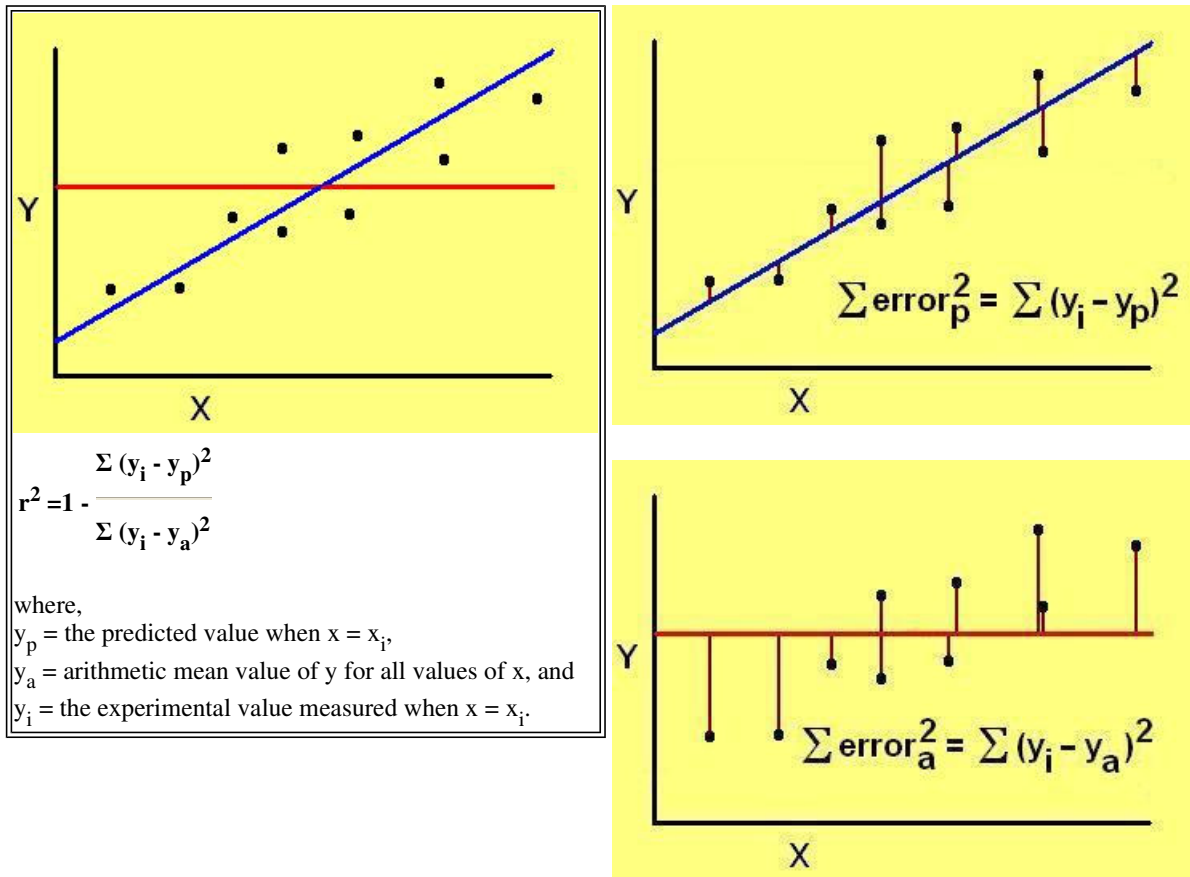
Note, however, that there is presently no definitive definition for ACF. Accordingly, the ACF parameter can be used as a relative weight conversion factor of the axes, instead of a unit conversion factor. Changing the value of ACF gives the user more flexibility on how to interpret the regression results. The ACF value is not a unique value. A large ACF value (>1) means that the y -axis values are weighed more strongly, and a better fit is observed for the data with high $[c]$ or $[S]$ values. A small ACF value (<1) means that the x -axis values are weighed more strongly, and a better fit is observed for the data with low $[c]$ or $[S]$ values.

In principle, modification of the ACF value can be used to highlight sensitivity of the regression results toward deviations in the x -axis or y -axis data values. No sensitivity is observed when the data and theory are perfect. In spite of this interesting observation with modified ACF values, I recommend that only one ACF value be used with your data, and that it be calculated similarly to the one example illustrated above that resulted in $ACF = 0.924$. If alternate ACF values cannot be defended, then they should not be used in the final presentation of the results. That is, I recommend that the ACF value be used only as a unit conversion factor, and thus remain true to the experimental propagation-of-error from the x -axis to the y -axis values collected and displayed on the original plot.

Correlation Coefficient (r^2)

When there is no pattern to a set of data, we can always determine the average value of the data. In this way, we'd be predicting too high half the time, and too low the rest of the time. In principle, we'll never be too far from the correct answer, and the error in our answer will be small. Without a pattern, the best we can do is $y = y_{\text{average}}$ for any value of x , where the probable error is expressed by the standard deviation (σ).

On the other hand, if there is a pattern, then we should be able to track it. The simplest pattern is a line, $y = mx + b$, where m = slope of the line, and b = y-intercept of the line when $x = 0$. We justify that the data follow a linear pattern by showing that using the line results in less error than using a simple arithmetic average. This is known as the goodness-of-fit of the line, and it is numerically expressed by the correlation coefficient (r^2):



If the predicted line has less error than the average value, then its distance from the measured value will usually be smaller than the distance between the average value and the measured value. Summed across all the points, it will definitely yield a smaller number, the numerator will be smaller than the denominator, and the value of r^2 will be close to 1.00. A perfect fit by a line will have $r^2 = 1.00$, whereas a poor fit by a line will have a low value. If $r^2 = 0$, then the predicted line is no better than a simple average of the data collected. With all linear predictions $r^2 \leq 1$, always.

This is easy enough to do by hand. It is even easier today with computers that do it for us.

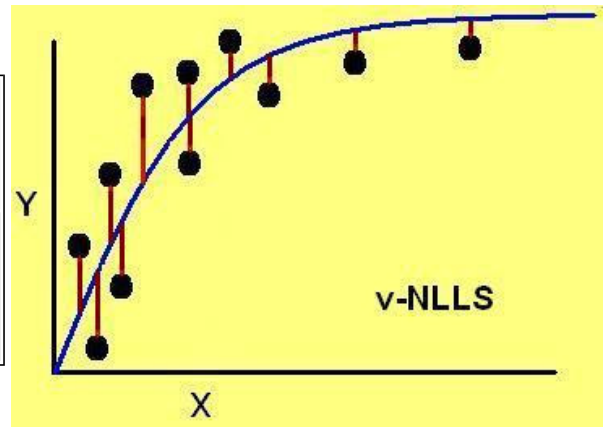
Correlation Coefficient (η^2)

The goodness-of-fit of a curve to a given set of data is based on how it compares with the goodness-of-fit of a simple horizontal line at the arithmetic average of the data. The goodness-of-fit of a curve to a set of data points is numerically expressed by the correlation coefficient (η^2).

The correlation coefficient η^2 coincides with r^2 in the linear regression setup. The eta-square term is used here for the correlation coefficient value of the untransformed data.

The η^2 assumes that a *vertical* minimum for the error terms corresponds to a better curve fit of the data collected. It is defined as:

$\eta^2 = 1 - \frac{\sum (y_i - y_p)^2}{\sum (y_i - y_a)^2}$
where, y_p = the predicted value when $x = x_i$, y_a = arithmetic mean value of y for all values of x , and y_i = the experimental value measured when $x = x_i$.



A perfect fit by a curve will have $\eta^2 = 1.00$, whereas a poor fit by a curve will have a low value. If $\eta^2 = 0$, then the predicted curve is no better than a simple average of the data collected. With all regressions, $\eta^2 \leq 1$, always.

This is not easy to do by hand. Computers are needed to perform these calculations easily and quickly.

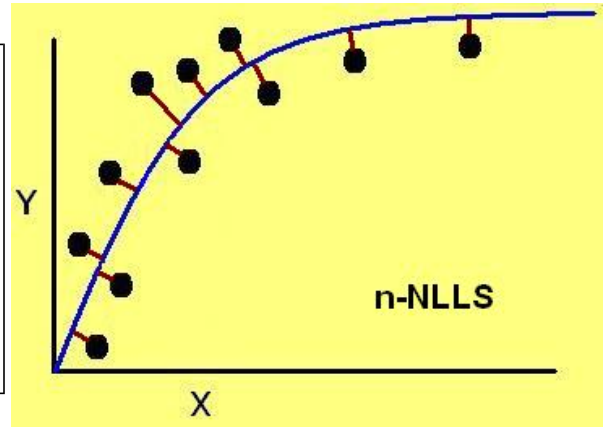
Correlation Coefficient (η_*^2)

The goodness-of-fit of a curve to a given set of data is based on how it compares with the goodness-of-fit of a simple horizontal line at the arithmetic average of the data. The goodness-of-fit of a curve to a set of data points is numerically expressed by the correlation coefficient (η_*^2).

The eta-star-square term is used here for the correlation coefficient value of the untransformed data.

The η_*^2 assumes that a *normal* minimum for the error terms corresponds to a better curve fit of the data collected. It is defined as:

$\eta_*^2 = 1 - \frac{\sum Q_i^2}{\sum (y_i - y_a)^2}$
where, Q_i = the normal distance of the data point to the nearest point on the predicted curve, y_a = arithmetic mean value of y for all values of x, and y_i = the experimental value measured when $x = x_i$.



A perfect fit by a curve will have $\eta_*^2 = 1.00$, whereas a poor fit by a curve will have a low value. If $\eta_*^2 = 0$, then the predicted curve is no better than a simple average of the data collected. With all regressions, $\eta_*^2 \leq 1$, always.

This is too difficult to do by hand. Computers are needed to perform these calculations easily and quickly.

Do's & Don't's

Do try to optimize your data using all of the regression methods available.

Do try to minimize data error and to use as many data points as possible. Try to have an even spread of the data on the plots, so as to have data present in low, medium and high concentration ranges.

Do be sure to use all of your original data for your final graphs, reports, and publications. Keep an eye on the "N = xx of xx" in the lower left corner of the graph. It is there to remind you if any data have been dropped for the current regression results being displayed. (N = number of data points being used out of the total number of data points available.)

Do present your conclusions about the goodness-of-fit of the data by showing the original untransformed data and graph with the optimized regression curve drawn. Showing the transformed data and graphs is to be strongly discouraged because it can be potentially very misleading. If a linear regression was used to optimize the equation parameters, it does not mean that the linear regression and its transformed data are more important than the original graph and its untransformed data.

The linear regression is just a mathematical tool, and the goodness-of-fit of the linear regression on the transformed data is typically not the objective. The objective is to work with the untransformed data and the original graph, and to see how well the optimized parabolic equation fits the data in this plot.

In other words, noting how the parameters were optimized is only important as a record of the methods used. It's an important footnote that must be told. However, the actual values of the optimized parameters plus information about how well they predict the untransformed data (based on a criteria that applies to the untransformed data) are typically the only two things that really matter.

Do try to locate areas that do not adhere well to the theoretical predictions. These areas may be easier to locate if you run several regression analyses, each with different data points selected for use in the regression runs.

Do report the goodness-of-fit of your optimized equation on the untransformed data set. Use either the η^2 or η_*^2 values for this.

Do compare the regression results among the various regression methods using only the η^2 or η_*^2 values.

Do not compare the η^2 value with the η_*^2 value (even if both are applied to the same graph) because they refer to different goodness-of-fit criteria.

Do not use r^2 to describe the goodness-of-fit of the original untransformed graphs. The r^2 correlation coefficient is for the transformed plots only. It describes the goodness-of-fit of the regression line on the transformed data.

Do not compare the r^2 of one regression method with the r^2 of another regression method. The r^2 value describes the goodness-of-fit of the regression line for a specific graph. If either the y-axis or x-axis units change, then the corresponding r^2 values of these two graphs cannot be compared. Such a comparison is nonsensical.

Do not compare the r^2 , η^2 or η_*^2 values of a given regression method. They each refer to different criteria for the goodness-of-fit.

Do not forget that this computer is just a machine. It does have some computational limitations. These limitations can be easily displayed when using very few data points in the lower left portion of the graph, especially when using just a few points that also do not fall on the typical path of a parabolic equation. Computer failure to do the math correctly nearly always results in a regression curve that is not anywhere near the middle of the cluster of data points used in the optimization procedure. When a regression analysis results in a predicted curve that does not pass through the center of the group of data points, *plus* the optimized parameter values are enormously large or small, then you know that the computational limits have been exceeded. If the optimized parameter values are indeed very large or very small, but the regression curve predicted remains near the middle of the cluster of data points used, then you are probably still well within the computational limits of your computer.

Do not rely too heavily on your conclusions and **do not** overstate your conclusions when your conclusions are based on just a few data points that were collected over a narrow range of conditions. Increase the range of your data, and get as much data as possible. If you have a cluster of data in one region of the graph, then your confidence of the results in that region of the graph will increase. However, if that region of the graph happens to correspond to a non-parabolic behavior in your particular experiment (that is, you have a slight theory error that is expressed in that region of the graph), then a large cluster of data in that region of the graph will affect the entire regression results. Therefore, **do** try to collect data evenly across all regions of the graph. The log-log and semilog graphs are especially useful for determining if the data are evenly collected across all regions of the graph.

Do not ignore data. It is very risky and sometimes unethical to do so. This LMMpro allows you to ignore data with just a click of the mouse in the *Data* window, but this practice is only to help you play with the data and thus evaluate various "what if" situations. These various "what if" situations, when used correctly, can help you locate areas on the graph that are displaying some kind of error, such as data error or theory error. You can then proceed with a plan to determine if problems exist with the data (such as precision and/or accuracy) or with the theory. In general, it will be difficult to detect any theory error if there is also some concern with the precision or accuracy of the experimental data values.

Welcome to LMMpro!

This program will help you optimize the equation parameters for the **Langmuir Equation** and the **Michaelis-Menten Equation**.

Tutorial: Overview of Software

Select a topic:

- [Demo Mode vs. Full Access Mode](#)
 - [Software Registration](#)
 - [Check for Updates](#)
 - [Options](#)
 - [Run Demo](#)
 - [Start Optimization Program](#)
 - [Data Entry](#)
 - [Regression Execution Errors](#)
 - [Graph Commands](#)
-

- [Tutorial: Overview of Theory](#)

<Reference = CTS-0>

Demo Mode versus Full Access Mode

Without registration of the software, the program will only run in ***Demo Mode***. This mode will allow you to run all of the features offered by the program, but you may only use the data sets provided. Running the program in this mode is free of charge, and the three demo data sets encoded in the program are intended to satisfy most teaching objectives about these equations, as well as about some of the limitations of the various regression methods commonly used. It is also recommended that you run this program in this mode prior to registering it so that you can be certain that it will function properly on your computer.

With registration of the software, the program will run in ***Full Access Mode***, which simply means that you can now also import your own data sets, create your own data sets, modify the data sets, and save the data sets. Registration of the software allows you to construct more complex data sets for your teaching objectives. The registration of the software will also allow you to apply these regression tools to your own research data sets and research projects.

Note that the registration of the software applies to only one computer per Serial Number entered. If the software is copied over to another computer, it will only run in *Demo Mode* in the other computer. The Serial Number and Registration Code obtained with the first computer will be recognized as an incorrect number-code combination in the second computer, which means that in the second computer it will run in *Demo Mode* only.

You may re-install the program into the same computer as often as you like. It will not affect the registration of the software. The only requirement is that it be reinstalled into the same computer where the original registration process took place. Also note that you must remember the Serial Number you were given when you first purchased the software through Alfisol, LLC. To return to *Full Access Mode*, you will also need to re-register the software every time you choose to re-install it. Re-registration is free and automatic with the click of a button; no purchase is necessary. If you do not remember the Serial Number, email technical support at www.alfisol.com for assistance. Note that the registration information will often survive the uninstall and re-installation procedure. If this occurs, then the needed codes will be found and your program will remain in *Full Access Mode*.

How to Register the Software

You must acquire a Serial Number in order to register the software. If you do not have a Serial Number, go to www.alfisol.com and get one by first purchasing it. Continue with the steps below only after you have acquired a Serial Number. You must be connected to the internet to complete the registration process. Registration of the software is a one time event.

Perform the following steps to register the software:

- Start the LMMpro program.
- Click the "program information" button.
- Click the "register now" button. Note that this button is deactivated if the program is already running in Full Access Mode. This occurs if the program is already registered.
- Enter the Serial Number in the space provided.
- Click the "submit" button.
- The program will automatically retrieve the "Registration Code" via the internet. Allow a little time for this process to complete.
- You will be prompted when the registration process is complete.
- If all went well, you will no longer need the internet to run the program.
- Click the "cancel" button to end your session with the Program Information window.
- Close all windows and restart LMMpro. From now on, LMMpro will always start in *Full Access Mode*, instead of *Demo Mode*.

Note that program registration is restricted to only one computer. **Do not** register from a computer that is not intended to hold the final resting place of the LMMpro program. Once registered, your Serial Number and the retrieved Registration Code will cause the software to work in *Full Access Mode* from only one computer, namely the one used during the registration procedure. Once registered, the Registration Code cannot be changed. You may, of course, purchase another Serial Number at any time, which will allow you to retrieve a completely new Registration Code. Once again, register the software while using the computer that also happens to be the final resting place of the software program.

How to Check for Software Updates

LMMpro provides a quick and easy way to see if a new version of LMMpro has been released that fixes any errors encountered. Note that your current Serial Number for this software is valid for all 1.xx versions of this software (that is, 1.00 to 1.99). No further purchases for a new Serial Number are required for upgrades involving versions 1.00 to 1.99. On the other hand, if running in *Demo Mode*, then no Serial Number is needed.

It is strongly recommended that you check for upgrades from time to time as these upgrades are free and these revised programs are intended to fix any coding errors encountered. Also, as a favor to your colleagues that also enjoy using LMMpro, please let Alfisol, LLC, know of any errors you may find in the program as soon as possible via email. To report a problem, click on the "Report a Bug" button, which takes you to the www.alfisol.com message web page. Thank you.

Perform the following steps to check for software updates:

- Start the LMMpro program.
- Click the "program information" button.
- Click the "check for updates" button.
- The program will automatically retrieve the latest version information via the internet. Allow a little time for this process to complete.
- The retrieved information is presented to you. If no updates are available, then close all windows. If an update is available, then go to www.alfisol.com for further instructions for your upgrade.

Prior to installing a new version, you must uninstall the old version. Typically, this will not erase your registration codes. Your old registration codes will apply to your new 1.xx version upgrades. However, just in case it does get erased, copy the registration information prior to the uninstall procedure. You will find this information by clicking on the "program information" button.

Options

The LMMpro options menu can be reached from the startup window or from the regression analysis window. This menu allows you to adjust how the program performs.

Mouse:

Under the mouse tab you will find four check boxes. They all refer to the tooltip information, which lights up when the mouse is passed over a datum point on the regression analysis window.

- **"On mouse over, ..."** : Check this box to activate the tooltip option. If this box is checked, then you can turn it off later by right clicking on the graphs and turning off the "Show Point Values". If this box is not checked, then the tooltip is always off.
- **"Original (x,y) data"** : Check this box to see the original data values when the mouse touches a datum point.
- **"Modified (x,y) data"** : Check this box to see the modified data. Note that the modified data is different with each regression analysis.
- **" $y = mx + b$ for linear regressions"** : Check this box to see the slope and intercept of the linear regression analysis shown in the graph when you touch any data in the graph.

Note that some of this information is also displayed in the lower edge of the regression analysis window when the keyboard arrows are used.

Colors:

This menu allows you to modify how the data are presented. You can also modify the shape and size of the points here. If modifications are made inside this tab, and if a regression analysis window is open, then the program will ask if you wish to redraw the regressions immediately with the new display choices made. If you decline, then the choices made will take effect on your next regression activity request.

- Normal = regular data or lines
 - Marked = marked data.
 - Highlighted = highlighted data or lines
 - Used = the data are not ignored in the regression analysis
 - Not Used = the data are ignored in the regression analysis
-

Regressions:

This menu allows you to decide which regression methods are to be calculated or not calculated. For those methods that require an iteration sequence, you are also given the option of controlling how many iterations to perform before it flags you for further instructions (namely, "continue regression for another set of iteration loops" or "cancel regression").

It is recommended that you uncheck those regressions that you know to be giving you problems. If a given regression fails, the program will automatically close down the display of that particular regression method. The LMMpro program is coded in such a way that it tries to keep it from crashing or freezing up. That is, if a regression fails, the computer hopefully does not also crash and freeze up. If the regression does not fail, but still results in a very strange answer, then the auto-scaling in the isotherm graphs may appear strange. After a few mishaps, you can uncheck the regression method that is causing the problems, and you thus avoid the constant nuisance of watching it fail or do strange things.

Typically, problems occur when the number of data points are too few to accurately optimize the regression method. A typical error results in negative numbers for the maximums or for the constants or for both. When negative numbers are obtained, the program will ask you whether or not you still wish to have the regression result displayed.

Start Up:

This menu allows you to select what to display on the start up window. It is intended to remove distracting buttons from the start up window (that is, the window that first appears when you start the program).

- "Tutorial" and "Run Demo" : You may wish to uncheck these once you are familiar with the program. Naturally, if you are not a registered user, or if you use this program for teaching purposes, you would probably wish to keep these buttons visible.
- "Registration Information" : There is probably no harm in leaving this button visible even after you are completely familiar with the program. It is through this button that you can get information about program updates.
- "Start Langmuir O.P." and "Start Michaelis-Menten O.P." : Typically, you would use one or the other. If so, then you can hide the start button of the one that you rarely use. Note, however, that at least one of these must be checked.
- "Restore All Options to Defaults" : As stated, click here to reset all options to the way it was when you first downloaded the software. This affects the options only. It does not affect the registration information.

Run Demo

LMMpro has three demo data sets that are intended to serve two purposes: (1) use as a teaching tool, and (2) use as a practice or demonstration tool. Its use as a teaching tool is most important. To run a demo, simply click on the "Run Demo" button on the start up window. If the button is not there, check the selections made on the start up tab under options, adjust as needed, and restart the program. If the program is not registered, then it is assumed that your only objective for this software is to use it as a basic teaching tool about the mathematics of various regression analyses, about the impact of data error, and about the impact of slight errors in the theory.

When you select "Run Demo", you will see a small window with options to select which demo to run. It may also ask you to select which type of example to run (Langmuir or Michaelis-Menten), but this question is hidden if the start up tab under options has only one run type pre-selected for default start up conditions.

The three demos shown are as follows:

- **Demo 1:** This is a very simple situation. The data were artificially selected so that a perfect fit would be found for $K = 0.0106$ and $\Gamma_{\max} = 838$, or $K_M = 94.34$ and $V_{\max} = 838$. The results are not affected by the number of data points available for the regression analysis.

Lesson: If the data are perfect, and the theory is perfect, then it does not matter how the regression is performed. All regression methods will reach the same conclusions. Much discussion can be pursued here about the nature of mathematics and the multiple paths allowed by mathematics to reach a similar conclusion.
- **Demo 2:** This is a slightly more complex situation. The data were artificially selected based on the same formulation applied to Demo 1. This time, however, the data were evenly increased or decreased by 10%. The regression results vary and this is especially true when some of the data are checked "not used" so as to slightly disturb the evenness of the error distribution.

Lesson: Although the theory may be perfect, the extraction of the correct values of the equation parameters can be difficult with "real world" data sets. This is particularly true if the data collected for the regression analysis are few in number. Much discussion can be pursued here about the primary cause of regression bias: *data error*. How is it recognized? How is it minimized?
- **Demo 3:** This is a more complex situation. The data are real. The data error is small, as noticed by the close reproducibility of the data. This time, however, there is probably a small error in the theory. For the most part, the theory seems correct and the slight error may go unnoticed.

Lesson: Although the data may have little scatter, and the data pattern may be parabolic, it does not necessarily mean that everything is perfect. We now see significant expressions of regression bias. Much discussion can be pursued here about the second cause of regression bias: *theory error*. How is it recognized? If present, then is it that the parameter values have changed, and, if so, how and where? Note that the

identification of how and where the parameters changed can be a strong contribution to science. Finally, if there is some theory error present, then should the theory be dropped, modified or simply noted as "almost correct"? If it is "almost correct", then is it ethical to drop the word "almost" in your reports or publications about the process studied? Review Schulthess & Dey (1996, *Soil Sci. Soc. Am. J.* 60:433-442) for a detailed discussion about this particular data set.

For the three data sets, the axes conversion factor (ACF) value is fixed at 0.924. If the program is registered, this value can be changed if so desired.

Clearly, if the program is registered, then you will be able to construct your own data sets, and thus challenge your students with more complex analytical problems. On the other hand, for most teaching environments and needs, these three demos will be sufficient. To be sure, with registration of the program you'd also be able to use the software for your own research needs.

<Reference = CTS-5>

Start Optimization Program (O.P.)

LMMpro can analyze two types of parabolic equations: the Langmuir Equation, and the Michaelis-Menten Equation.

If your program is registered, then you will be able to run the program using your own data. If not, then the two buttons that perform this function are temporarily hidden and replaced with the words "Demo Mode".

To run the program, simply click on "Start Langmuir O.P." or "Start Michaelis-Menten O.P."

Note that if you usually use only one of these optimization programs, you can choose to hide the button that activates the optimization program of the one you do not use. Click on "Options" button & "Start Up" tab to set this up according to your preferences.

<Reference = CTS-6>

Data Entry

You reach the Data Entry window by clicking on the "Start Langmuir O.P." or "Start Michaelis-Menten O.P." buttons on the startup window. You can also reach a modified version of this Data Entry window by clicking the "Edit" button in the Data Display window. The software must be registered if you wish to access the Data Entry window.

There are numerous things that can be done in the Data Entry window. These are tabulated below:

Action:	Effect:	Comments:
Data entry in data rows.	Numbers are filled in.	Be sure to have more than one row of data entered.
Text entry in units boxes.	Data units are recorded.	These units are not used in the regressions. Its purpose is only for clarity to the user about the data values entered.
Text entry in comments box.	Comments about the data set are recorded.	The comments entered are not used in the regressions. Its purpose is only for clarity to the user about the data entered.
Click on row number	Row is highlighted.	This allows you to delete the row. Press <delete> key to delete the data shown in row. Caution: There is no undo option for this action.
Click on <i>x</i> -axis column header (labeled c or S).	Numerically rearranges the data in ascending order.	Program uses the <i>y</i> -axis data as a tie-breaker. Data order does not affect the regressions. Its purpose is only for clarity to the user about the data entered.
<Save Data>	*.LMM data file is created & saved.	If filename already exists, prompt for "save as..." or "overwrite" appears. Data must exist for this action to work.
<Import Data>	*.LMM data file is imported.	Data Entry window will close. Data Display window will open after data import is completed.
<Run Regressions>	Program executes the regression analysis.	Window will close. Regression Results window will open. Review the Options window to control how the regressions are performed and presented.
<Clear Data>	All data displayed are erased.	Caution: There is no undo option for this action.
<Cancel>	Close this window.	No records are kept of modifications made.

<Usage>	Click on column header = all boxes are on (checked) or off (unchecked). Checked box = data in row are used. Unchecked box = data in row are ignored.	If data are ignored, then regression assumes that they do not exist. Caution: The all-on or all-off column header command will not apply to a cell that is currently highlighted.
<Marked>	Click on column header = all boxes are turned off (unchecked). Checked box = use special symbols when displaying this point on the graphs. Unchecked box = use regular symbols when displaying this point on the graphs.	Marking the point does not affect the regression. Its purpose is only to give the user more control over how to draw the graphs, or to emphasize a particular subset of the data in the graphs. Caution: The mark-all-off column header command will not apply to a cell that is currently highlighted. There is no mark-all-on command for this function.
Fix K or K_M	Checked box = Run regression using fixed value entered. Unchecked box = Ignore value shown (if any).	This forces the regressions to find the best results with this parameter fixed. If fixed, the word "fixed" appears on the results table on the top of the column for this parameter.
Fix $\Gamma_{\max}$ or $V_{\max}$	Checked box = Run regression using values shown. Unchecked box = Ignore values shown (if any).	This forces the regressions to find the best results with this parameter fixed. If fixed, the word "fixed" appears on the results table on the top of the column for this parameter.
ACF	Enter the Axes Conversion Factor	This value is required for the n-NLLS regression. Default = 1.

Regression Execution Errors

Regression methods work best when there is plenty of data to work with. Common errors include:

- **Regression returns extremely high maximum values for $\Gamma_{\max}$ or $V_{\max}$:**

This can occur if all of the data are clustered around the lower left corner of the graph. If none of the data are in the mid-range or higher portions of the graph, then a small error in the data can easily result in an unusually sharp rise and an unusually large maximum. Typically, this problem will also cause the software to exceed the maximum number of iterations allowed.

Fixing the problem: Get more data using higher concentration (c) or substrate (S) values. This error will also affect the display of some of the graphs, which can be fixed by manually rescaling the graphs.

- **Regression returns negative parameter values:**

This can occur if all of the data fall near the lower left corner of the graph and if the modified data plotted on the linear regressions results in an inverted slope. In addition to an inverted slope, there can be a y-axis intercept in the linear regression that is on the wrong side of the regression graph. Technically, the regression is completely mathematically correct, and the "best" parameter values have been found. Nevertheless, negative parameter values are not logical answers.

Fixing the problem: Get more data using higher concentrations (c) or substrate (S) values. When this problem occurs, LMMpro will automatically inform you that the results are nonsensical and ask you if you wish the program to continue or to cancel the particular regression causing the problem. This error will also affect the display of some of the graphs, which can be fixed by manually rescaling the graphs or by instructing LMMpro to cancel the particular regression(s) causing the problem.

- **Value of data is greater than estimated maximum value ($\theta \geq 1$):**

This problem occurs if the scatter in the data is such that a datum point exists that happens to be higher than the maximum value estimated by the regression method. This error applies only to the log-log regression method, where the estimated θ value must be less than 1.

Fixing the problem: LMMpro will automatically inform you of this error and it will also cancel the log-log regression when this occurs. For subsequent regression runs for log-log, uncheck the "usage" box in the Edit Data window for this high datum point. However, this will also affect the results obtained with all the other regression methods. Accordingly, this modified data set is probably only useful for a study of the log-log regression results. An alternate solution is to deactivate the log-log regressions (see "Options" button / "Regressions" tab).

Also note that placing data in the "not used" category, may cause the high "not used" data to be hidden from view in the log-log graphics. This is probably because the 0 value of those high "not used" data points resulted in the log of a negative number, which is not allowed. If you use the arrows to highlight the data points, but a "not used" data point is not highlighted in the log-log regression, then it is probably because this kind of error has occurred and the point was dropped from the display. This situation only applies to "not used" data. With "used" data, this situation will result in an error message and the closing of the log-log regression analysis.

- **NaN in regression results table:**

NaN = not a number. No number was generated. Perhaps the regression was automatically closed.

Fixing the problem: Check that enough data are used in the regressions.

Graph Commands

There are numerous commands that can be used on the window that displays all of the results. These are tabulated below:

Action:	Effect:	Comments:
Tool bar buttons.	Opens various windows (Data Edit, Options, Help).	These are explained elsewhere.
[X] on upper right corner of an internal graphics or table box.	Hide the graphics or table box.	Its purpose is to give the user more control over what is shown on the monitor.
Selecting an internal graphics box.	If hidden, show the graphics or table box. If not hidden, show the graphics box in a new, larger window.	Click on the name of the regression above the graph. Its purpose is to allow larger or even full-screen display of any graph.
Mouse rollover data	Tooltip shows information about the data and regression.	For tooltip to work, it must be allowed in the Options window. It can be turned off by a right-click anywhere inside a graph and deselecting "Show Point Values". To modify the information presented by the tooltip, go to Options window and Mouse tab.
On results table, click on regression name.	Row is highlighted. Corresponding regression lines are also highlighted on graphs.	Its purpose is to help user identify which line is which.
Keyboard arrows: <up> <down>	Next row (up or down) of results table is highlighted.	Corresponding regression lines are also highlighted on graphs. Its purpose is to help user identify which line is which.
Keyboard arrows: <left> <right>	Data are highlighted in the order entered.	Note that the same datum point is highlighted on all the graphs. The original values of each datum point is displayed on the lower bar. Its purpose is to help the user determine which point is which.

Area-Box Trace in Graphics (click-hold-drag mouse)	Area traced becomes area shown by the graph.	To undo, right click on graph, and select "Un-Zoom" (will undo last zoom modification), "Undo All Zoom" (will return to first scaling options used, and it is not the same as the default scale), "Set Scale to Default" (note that default scaling options may include extra white space), or "Scale Options" (will let user define the scale parameters). Its purpose is to help the user control the graph displayed.
Turning the Mouse Wheel over a particular graph	Auto zoom in or out for that particular graph.	Its purpose is to help the user control the graph displayed.

<Reference = CTS-9>